THREE ESSAYS ON INVESTMENT IN HUMAN CAPITAL

By

SHAH DANYAL

A Dissertation Submitted to the Graduate School at Middle Tennessee State University in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSPHY- ECONOMICS

Murfreesboro, TN
May 2010

Approval Page

THREE ESSAYS ON INVESTMENT IN HUMAN CAPITAL

By

SHAH DANYAL

A Dissertation Submitted to the Graduate School at Middle Tennessee State University in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSPHY- ECONOMICS

Murfreesboro, TN
May 2010

Approved by:

Professor Bichaka Fayissa, Committee Chair

Dr. Anthon Eff, Committee Member

Dr. Gregory Givens, Committee Member

Dr. Franklin A. Michello, Committee Member

Dr. Charles L. Baum, Department Chair, Economics and Finance

Dr. Michael D. Allen, Dean, College of Graduate Studies

TO MY WIFE

SYEDA NABEELA NAQVI

ACKNOWLEDGMENT

I would like to thank my dissertation committee: Bichaka Fayissa (chair), E. Anthon Eff, Greg Givens, Franklin A. Michello. I would also like to thank Charles Baum and Joachim Zietz for providing me an initiative in conducting research. Furthermore, I am grateful for the support of Saeed Chowdhury, Wei Kang, Pankaj Maskara, and my loving family: Annaheeta, Shikoh, Shariq, Ali Raza, and Nabeela.

ABSTRACT

This dissertation deals with three interrelated essays on investment in human capital which are covered in chapters two, three, and four. In the purview of the human capital theory, prominent economists have addressed the role of new and better skills in creating job opportunities for workers, not only to find and hold on to jobs, but also to improve their living standards through higher earnings by upgrading their skills. Many researchers have also addressed the role of education on health and lifestyle choices with mixed findings.

In chapter two, we investigate the "Impact of Computer Skills on Wages" in the U.S. using NLSY79 panel data set, staggered every two years from 2000-2006 for a cross-section of 12,686 individuals. Specifically, the essay examines the controversy in the literature whether there is a wage premium due to the acquirement of computer skills by individuals confirming the skill biased technological change (SBTC) hypothesis. By defining computer skills as having a computer with Microsoft Windows or NT, at home and using the fixed-effects model and the instrumental variable technique, the study finds that individuals possessing computer skill do, indeed, earn a wage premium, confirming the SBTC hypothesis.

Chapter three titled "Effects of Education on Health: A Panel Data Study from NLSY" investigates the effect of educational attainment on the individual's health status as measured by the inability to work for health reasons. Based on the unique data set and the Arellano-Bond estimation methodology, the study finds that educational attainment has a

positive effect on the quality of an individual's health status. The chapter also bridges the gap in the literature by using the robust fixed-effects model and Arellano-Bond to analyze the impact of education on the health status after controlling for the unobserved individual heterogeneity and the endogeneity problem arising from the interaction between education and the measure of the health status.

The third essay, "Impact of Education on Lifestyle Choices: A Panel Data Study from NLSY79," examines the effect of education on different lifestyle variables using NLSY79 panels for 1992 1994 and 1998 in chapter 4. Using smoking, drinking, marijuana use, and cocaine use as lifestyle variables, the study addresses the joint determination of lifestyle variables within the framework of Seemingly Unrelated Regression (SUR) model. After controlling for the unobserved individuals heterogeneity by robust fixed-effects model extended to SUR model, the study finds that educational attainment does not necessarily have a significant effect on lifestyle choices. While future study with adequate data base and alternative methodology may find different results and explanations, perhaps, the finding of this essay suggests that it is the health knowledge that affects lifestyle choices (such as warning labels on cigarettes, alcohol products, and nutritional contents on processed foods) rather than the educational of individuals. The marginal contribution of this essay to literature is the use of the robust fixed-effect model in context of SUR model to analyze the impact of the cross and within correlations of educational attainment on the lifestyle choices.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Although Adam Smith (1776) did not coin the phrase, the concept of "human capital" has evolved since the advent of supply and demand. Adam Smith defined labor in relation to human capital thus: "The acquisition of such talents, by the maintenance of acquirer during education, study, or apprenticeship, always costs a real expense which is a capital fixed and realized, as it were, in his person. Those talents, as they make a part of his fortune, so do they likewise that of the society to which belongs. The improved dexterity of a workman may be considered in the same light as a machine or instrument of trade which facilitates and abridges labor, and which, though it costs a certain expense, repay that expense with a profit." It was Pigou (1928, 27-28) who coined the term "human capital" and denoted the difference between the investment in human capital and the investment in material capital. The understanding was further elaborated by Mincer (1958), who concluded that symmetrical income distribution is due to the difference in investment in human capital. However, Becker (1964) found that "education, training, and health are the most important investments in human capital." According to Becker, human capital is like the physical means of production, as it increases productivity. Schultz (1980) believes that human capital distinguishes developed from underdeveloped economies: developed economies have a greater stock of human capital and therefore better education and health. He is of the opinion that economies' expenditures on education and health should be treated as investment rather consumption expenditures.

Considering human capital as consisting of education, skills, and maintaining good health, the present three essays deal with the subject of the investment in human capital. All three essays address the role of education and skills. The first essay, "The Effect of Computer Skills on Wages," falls in the category of investment in individuals' education and skills. The second and thirds essays, "Impact of Education on Health" and "Impact of Education on Lifestyles," incorporate the broad category of investment in health. In all three essays, however, econometric techniques are engaged to control for individual specific heterogeneity and where possible the causal effects of education through elaboration of the dependent variables. In the first essay, the dependent variable is wages; in the second, health status; and in the third, multi-lifestyle variables. On these topics, a substantial amount of research is available in the literature, but there is a dispute over the unobserved heterogeneity or individual specific heterogeneity. The uniqueness of the three essays is that the unobserved heterogeneity problem is treated differently than in the existing literature.

There has been a significant amount of research relevant to the premium paid on computer skills and on health topics using instrument variables technique for controlling individual specific heterogeneity factors. The instrument variables technique assumes that if correctly applied it will filter out non-causal effects of education or skills. The earlier studies generally used the instrument relation with family background. These instruments are questioned (Grossman, 2005). The latter studies used institutional characteristics for instrument variables. However, the studies relevant to the subjects of the present three essays using the instrument variable technique face the same general criticism of using the method of instrumental variables, based on two conditions for good instrument variables. One condition is the exclusion restriction, and the other is known as "weak instruments." The exclusion restriction assumes that the instrument variable should be correlated with the explanatory variable that has an endogeneity problem and should not be a direct determinant of the dependent variable. The weak instruments problem is that there should not be low correlation between the explanatory variables considered having endogeneity (Stagier and Stock, 1977; Bound et al., 1995). The approach in these three essays is to address unobserved heterogeneity, also known as endogeneity, due to individual specific heterogeneity by employing the fixed-effects model using panel data. This still leaves the question of endogeneity due to reverse causation. In the first essay, chapter 2 of this dissertation, unobserved heterogeneity is controlled by the fixed-effects model. To control for both kinds of endogeneity, the instrument variable technique is applied, although admittedly the instrument variable is vulnerable to criticism as stated above. The paucity of panels did not lead us to use the Arellano-Bond estimates, which addresses both endogeneity problems. In chapter 3, both the fixed-effects model and the Arellano-Bond technique are incorporated, while in the fourth chapter only endogeneity due to individual specific effects is addressed.

The data interpreted for all three essays are obtained from the National Longitudinal Survey of Youth 1979 (NLSY79). The NLSY79 includes youth aged 14 to 21 as of Decmber 31, 1978. Initially, the survey comprised 12, 686 individuals, who were interviewed annually through 1994 and on a biennial basis since 1996. All the essays use a different set of panel years, so questions relevant to the studies should be exact in content and comparable. Over time, the retention of individuals was different in different years. The decrease in the number of respondents is due to attrition. The data used in the present studies consider some observations as missing so they do not create a systematic bias.

The 1980s witnessed substantially increased wages of college-educated individuals since the 1930s. Murphy and Welch (1991) found that the premium for college education is 58 percent. One explanation given is that great innovation in technology would favor highly educated people. The dominant technology innovation is in the computer sector, and educated people have adapted more to computer use and are rewarded with a great premium. This is known as the Skill Biased Technology Change (SBTC) hypothesis. The first essay explores this hypothesis and finds, through the use of the robust fixed-effects model and instrument variable technique, that there is a premium on computer skills.

There is a pecuniary return on education. The question then arises whether there is a health return on education. Grossman's (1972) model of demand for health reveals that schooling plays a causal role by increasing the efficiency of the household production function of health. The second essay explores Grossman's hypothesis by applying the robust fixed-effects model for controlling unobserved heterogeneity and finds that schooling has an insignificant effect on health. However, when the Arellano-Bond technique is applied, it is deduced that education is a significant variable in determining health status.

The third essay explores the following: if education has a causal role in increasing the efficiency of the household health production function, then education will affect the various behavior or lifestyles variables, which are the input to the production function of health. The essay addresses the joint determination of multi-lifestyle variables, using the Seemingly Unrelated Regressions (SUR) model. By using SUR, the third essay contributes to the existing empirical literature. The general SUR model assumes that there is a cross correlation among the equations and no within correlation. Within correlation is relaxed by introducing the assumption that there is within correlation through individual specific errors. The third essay extends the fixed-effects model into SUR, an approach not found in the existing literature. The finding is that education plays no significant role in explaining the lifestyle variables selected for the essay.

All three essays were written independently, so the explanations common among them are not merely repetition. The essays explain the application of the robust fixed-effects model, considered new to the subject area of each essay. The conclusion derived from these essays can provide guidelines to academic institutions and policymakers. Academic institutions may offer subjects that enhance students' acquisition of technology skills and concurrently administer campus rules prohibiting students' involvement in lifestyles injurious to health. The guideline for policymakers at the government level is to invest in education complementary to the development of technological skills, make and strictly implement policies that discourage lifestyles hazardous to health, and disseminate health information throughout the population.

CHAPTER 2

IMPACT OF COMPUTER SKILLS ON WAGES IN USA

Abstract

Using the U.S. NLSY79 panel data set, staggered every two years from 2000-2006 for a cross-section of 12,686 individuals, we investigate the effect of computer skills on wages. We define computer skills as having a personal computer with Microsoft Windows at home. We use fixed-effects and instrumental variable (IV) estimation techniques to investigate the effect of computer skills on wages. The results suggest that individuals possessing computer skills earn a wage premium, confirming the skill-biased technological change hypothesis (SBTC).

Keywords: wages, computers, technological change, education

JEL: I2, J31, O3

Impact of Computer Skills on Wages in the U.S.A.

2.1. Introduction

The association of wages and education in the eighties is widely agreed upon by most labor economists. The disagreement has remained with respect to the causation of the association (Autor et al., 1998; Bound and Johnson, 1992). One body of literature shows that the causation is due to technological change. This explanation rests on the assumption that innovation in technology favors better-educated people, causing their demand to increase (Krueger, 1993; Mincer, 1991). A large portion of technological change is brought about by computers, and workers with more schooling are likely to use computers at the workplace. Thus, computer technology has become complementary to human capital (Autor et al., 1998; Autor et al., 2003; Krueger, 1993). The notion that new technology has caused an increase in the demand for highly skilled workers, thereby increasing their earnings, is known as the skill-biased technological change (SBTC) hypothesis (Krueger, 1993). While Krueger's conclusion of the relationship between computer skills and wages continues to inspire new literature, the SBTC hypothesis has also drawn criticism and disagreement among researchers (Card and DiNardo, 2002; Lemieux, 2006).

Two main points in the SBTC literature are currently debated. One is the definition of computer skills, while the other relates to unobserved heterogeneity ignored in the analysis, which may lead to estimation bias on the relative impact of computer skills on wages for U.S. workers (Tashiro, 2004). This paper has two objectives. First, we define computer skills as the possession of a personal computer at home with Microsoft Windows or NT. According to *PC* magazine, more than 90 percent of personal computers use Microsoft Windows (AFP, 2009). This study attempts to verify the impact of computer skills on wages based on the National Longitudinal Survey of Youth 1979 (NLSY79) data for years 2000, 2002, 2004, and 2006. Second, most of the previous studies for the U.S. use cross-sectional data, unlike the present study, which employs panel data that allow us to account directly for individual-specific characteristics using the fixed-effects and random-effects models. To the best of our knowledge, none of the published studies for the U.S. incorporate the fixed-effects model for controlling unobserved individual heterogeneity.

Broadly speaking, endogeneity is of two types. One type is due to unobserved individual-specific effects, while the other may be due to possible reverse causation. Endogeneity due to unobserved heterogeneity is controlled by the fixed-effects model, while the present analysis also applies the instrumental variables (IV) technique to address the issue of reverse causation. We test for heteroscedasticity using the Breusch-Pagan approach and present a robust analysis. Our study follows the Krueger (1993) methodology and confirms the presence of SBTC hypothesis after controlling for unobserved heterogeneity and simultaneity bias between wages and computer skills. The rest of the paper is organized as follows. The next section reviews the related literature. Section III describes the data used for the study. The empirical methodology and results are presented in section IV. Finally, the conclusion summarizes the results.

2.2 Review of Literature

Many empirical studies have found significant results showing that the underlying cause of the observed pattern of wage inequality is the higher demand for skilled labor arising from technological change (Bound and Johnson, 1992; Katz and Murphy, 1992). The early studies did not apply technological change directly but concluded that the residual trend in the measured skill premium is due to the technology factor. Krueger (1993), however, studied the direct effect of computer use on wages using Current Population Survey (CPS) data for 1984 and 1989 and found that workers using computers earned about a 15 percent to 20 percent wage premium. Autor et al. (1998) later broadened the scope of the study by incorporating inter-industry factors and found that computers, capital, and college educated workers are complementary. They also confirmed that there is a wage premium for workers using computers. Daldy and Gibson (2003) and Dolton and Makepeace (2004) found similar results for New Zealand and the United Kingdom, respectively. They concluded that there is a substantial skill premium for workers using computers. DiNardo and Pischke (1997), however, raised doubts about whether the measured wage differentials are due to computer use. They found that workers using calculators, telephones, pens, and pencils show similar results on wage differentials. Hence, computer skills may not have an independent causal effect on wages. There may be unobserved heterogeneity factors in workers' productivity causing higher wages for individuals who use computers. Surprisingly, using a more recent wave of data employed by DiNardo and Pischke (1997), Spitz-Oener (2007) found that there is a significant return on computer use but no significant return to calculators, telephones, pens, and pencils. Krashinsky (2004), on the other hand, found that computers skills are not important casual determinants of earnings.

In an effort to find the effects of new technology on wages and employment in France, Entrof et al. (1999) observed that the total return to computer use is about 2 percent after controlling for unobserved heterogeneity. Doms et al. (1997) found that firms using new automation technologies such as programmable controllers, computer-automated designs, and numerically controlled machines have effects on wages in cross-sectional data at the plant level. However, longitudinal data analysis showed little correlation between skill grades and the adaptation of new technologies. Weinberg (2000) argued that one of the causes of the increase in women's relative wage is the ease of their adaptation of new technologies. New technologies allow computer use in the workplace, and female workers are relatively more likely to benefit from such jobs.

Dolton and Makepeace (2004) argued that if the estimates are large and consistent, as their study confirmed, then it is sufficient evidence of the wage premium for using computers. Although Krueger (1993) accounted for the heterogeneity factor indirectly, he did not control for the wage premium arising from the use of computers by applying the instrumental variables technique. Pabilonia and Zoghi (2005) used instrumental variables for controlling unobserved heterogeneity in their study for Canada and found that there is a positive and significant wage premium due to computer skills. Maskara et al. (2006) used NLSY79 data for the years 2002 and 2004 and found a significant wage premium for people with access to computer technologies at home. However, they did not control for the unobserved heterogeneity factor by using the fixed-effects model. This study bridges the gap in the literature with respect to the impact of SBTC in explaining the wage differences among U.S. workers by using the robust fixed-effects and instrumental variable empirical methodologies to determine whether or not computer skills result in a wage premium across heterogeneous individuals in the sample.

2.3. Data

Our empirical analysis is based on data from NLSY79, which is a nationally representative sample of 12,686 young men and women who were 14-22 years of age when they were first surveyed in 1979. These individuals were interviewed annually from 1979 through 1994 and have been interviewed on a biennial basis since 1996. We, however, use data only for the years 2000, 2002, 2004, and 2006 because the relevant survey question for computer skills was first introduced in 2000. For all the variables used in this study, the content of the relevant questions in the survey remain the same and are exactly comparable over the sample period. The original survey of NLSY79 started with 12,686 respondents, which decreased to 7,764 by the end of 2006 due to attrition; the difference is considered missing information. Unanswered questions are also considered as missing data. In the 2000 survey, the respondents' age ranged from 35 to 43. We use the survey question *"Do you have a personal computer at home running Microsoft Windows 95/98 or NT?"* to generate a proxy for computer skills.

The IV chosen is based on the survey question *"While working at [name of employer], did you receive any informal on-the-job training by making use of any self-study material or self-instructional packages such as manuals, workbooks, or computer-assisted teaching programs?"* The rationale for choosing this variable is that when the production process changes in firms, the workers go through training and are introduced to new technologies. This is likely to result in the acquisition of computer skills.

Table 1 gives the definitions of the variables used in the study, and the summary statistics are presented in Table 2. The dependent variable (average hourly wage) is a continuous variable, but all of the regressors are dummy variables. The data suggest that the average hourly pay in 2000 was $16.73 and it grew to $20.31 in 2006 (an increase of 20 percent). The standard deviation of hourly wages was $31.58 in 2006 as compared to around $20 in previous years. We find that the number of respondents reporting having access to a computer at home increased steadily from 58.8 percent in 2000 to over 82 percent in 2006. About 3 percent of the respondents had eight or fewer years of education, and 57 percent of the respondents had some level of high school education. About 36 percent had some level of undergraduate education. Almost 50 percent of the respondents were male, 25 percent were Black, 15.7 percent were Hispanic, and about 58 percent were married. Around 17 percent of the respondents were members of a workers' union, and more than 70 percent of them lived in urban locations. In terms of the regional distribution of respondents, 16 percent live in the Northeast, 23 percent in the North Central, and 42 percent in the Southern region of U.S.

In Figure 1, we show the trend of PC ownership at home based on the level of education. The graph shows that in 2000 a large proportion of people with a PC at home had college education, but by 2006 persons with only a high school education were the majority of home PC owners. During our sample period, PC ownership at home increased steadily for every segment, except for the group with fewer than eight years of education. In Figure 2, we separate the trend of PC ownership based on gender. The data suggest that home PC ownership has steadily increased over the sample period across gender, but females had higher PC ownership than males.

2.4. Methodology and Results

The baseline model for the impact of computer skills on wages in this essay is similar to Krueger's (1993) model, except that the present study uses panel data and controls for individual-specific effects. The dependent variable is the log of the hourly rate of pay. The explanatory variables include proxies for computer skills and several personal characteristics. The regression has the following functional form:

$$lnWAG_{it} = \alpha_i + \gamma COM_{it} + \beta X_{it} + \varepsilon_{it}$$

where WAG_{it} is the wage of individual i at time t. α_i is a dummy to account for individual-specific effects; COM_{it} is the computer skills of individual i at time t set as 1 for having a computer at home running Microsoft Windows 95/98,or NT, 0 otherwise. X_{it} denotes the control variables such as gender, education categories, race, marital status, family size, location of residence (whether metropolitan area or not; broad geographical regions such as northeast, central, south, and urban or rural), and union membership. ε_{it} is the disturbance term, which accounts for omitted factors and other random errors.

In Table 3, we present robust OLS estimates for the year 2000 and for our full sample period (2000-2006). For the year 2000, we get a significantly positive estimate (0.238) for the coefficient of COM, which is comparable to the Maskara et al. (2006) estimate (0.230) over the same period. However, their education variable is continuous, whereas our education variables are dummies for middle school, high school, and college education. The robust estimates suggest that people with a PC at home earn a wage premium of about 29 percent over their peers who do not possess computer skills. Married people earn about 10 percent more than their counterparts. These estimates are in line with those of previous studies (Maskara et al., 2006). For the full sample period, the coefficient estimate for the *COM* variable increases from 22.7 percent to 29.2 percent. However, the returns to computer skills may be suspect because of endogeneity due to unobserved heterogeneity or reverse causation.

After controlling for individual-specific effects, we find that there is a wage premium of 4.5 percent due to computer skills (Table 4). The results of the instrumental variable technique, however, suggest the wage premium due to computer skills is substantially high (121.7percent), perhaps due to difficulties associated with finding the right instruments as documented in the literature (Bound et al., 1995; Stagier and Stock, 1997).

There are two sources of unobserved heterogeneity or individual-specific effects. One source of individual-specific effects is random, while the other is fixed. Time-related unobserved heterogeneity is controlled by the random-effects model and between estimators. Individual fixed heterogeneity is controlled by the fixed-effects Generalized Least Squares Method (FEGLS). The results of the fixed-effects, random-effects, between-effects, and IV estimates are reported in Table 4. The Hausman test of difference between the fixed- and random-effect estimates rejects the random-effects model in favor of the fixed-effects model (Table 4) on which we base our discussion. The results indicate that computer skills have a positive and statistically significant effect on wages in all models. The coefficient estimate of computer skills is highest in the IV model and lowest for the fixed-effects model. On the other hand, the estimates for 0-8 years and 9-12 years of education are found to have a negative and statistically significant effect on wages in the random-effects model and between estimators. However, the coefficient estimates for 0-8 and 9-12 years of education is positive but insignificant, while the coefficient estimates for 13-16 years of education is positive and significant in the fixed-effects and IV models.

The random-effects and between-estimator models show significant and larger coefficients for computer skills because their impacts are not dampened by omitted individual-specific variables. The coefficient of computer skills in the fixed-effects model is 0.045, which implies that acquiring computer skills will increase the wage premium by 4.5 percent. Furthermore, our results do not provide evidence that union membership increases wages, while individuals living in urban locations tend to have higher wages.

2.5. Conclusion

This paper examines the question of causation and association between computer skills and wages of individuals. Previous studies related to the impact of computer skills on wages for the U.S used the IV methodology to address the issue of endogeneity arising from unobserved heterogeneity (Maskara et al., 2006). In this study, we employ a variety of models to deal with the problem of unobserved heterogeneity. We apply the robust fixed-effects model and IV methodology to address problems due to unobserved heterogeneity and endogeneity, respectively.

Krueger's (1993) study finds a positive and significant effect of computer skills on wages. After controlling for heterogeneity using the more robust fixed-effects model, this paper also finds a positive and statistically significant wage premium to computer skills acquirement. We also find positive effects of computer skills on wages in the IV model with very high return (121.7 percent), while the robust fixed-effects model yields a modest wage premium of only 4.5 percent, which is less than the 18 percent premium obtained by Krueger (1993). Thus the positive wage premium due to the acquisition of computer skills confirms the SBTC hypothesis. The policy implication of the study is that government incentives (such as tax credits) can be used to promote workers' skill upgrades and increase economy-wide productivity.

Table 1
Definition of Variables

Variables	Variable type	Variables Definition
WAG	*Continuous*	Hourly pay
COM	*Dummy*	1 for having personal computer at home running Microsoft Windows or NT, otherwise 0 (proxy for computer skills)
EDU1	*Dummy*	1 for education level of 0-8 years, otherwise 0
EDU2	*Dummy*	1 for education level of 9-12 years, otherwise 0
EDU3	*Dummy*	1 for education level of 13-16 years, otherwise 0
MST	*Dummy*	1 for being "Married," otherwise 0
GEN	*Dummy*	1 if Male, otherwise 0
BLK	*Dummy*	1 if Black, otherwise 0
HIS	*Dummy*	1 if Hispanic, otherwise 0
UNI	*Dummy*	1 if member of Workers Union, otherwise 0
URB	*Dummy*	1 if lives in Urban area, otherwise 0
SMSA	*Dummy*	1 if lives in Standard Metropolitan Statistical Area, otherwise 0
REG1	*Dummy*	1 if lives in region 1 (Northeast), otherwise 0
REG2	*Dummy*	1 if lives in region 2 (North central), otherwise 0
REG3	*Dummy*	1 if lives in region 3 (South), otherwise 0

Table 2
Descriptive Statistics

	2000			2002		2004		2006	
Variables	Mean	Standard Deviation	N	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
WAG	16.730	21.280	7,043	17.950	19.450	18.450	19.460	20.310	31.580
COM	0.588	0.492	7,985	0.712	0.452	0.783	0.441	0.821	0.382
EDU1	0.032	0.177	8,033	0.030	0.171	0.028	0.167	0.029	0.171
EDU2	0.588	0.492	8,033	0.572	0.494	0.567	0.495	0.562	0.496
EDU3	0.351	0.477	8,033	0.359	0.492	0.360	0.480	0.359	0.479
GEN	0.495	0.499	12,686	0.495	0.499	0.495	0.499	0.495	0.499
MST	0.580	0.493	8,031	0.586	0.492	0.581	0.493	0.571	0.494
BLK	0.250	0.433	12,686	0.250	0.433	0.250	0.433	0.250	0.433
HIS	0.157	0.364	12,686	0.157	0.364	0.157	0.364	0.157	0.364
UNI	0.168	0.374	6,646	0.170	0.376	0.178	0.382	0.174	0.379
URB	0.729	0.443	7,781	0.755	0.429	0.751	0.432	0.716	0.450
SMSA	0.936	0.244	7,928	0.812	0.390	0.815	0.388	0.929	0.256
REG1	0.156	0.363	7,973	0.155	0.362	0.153	0.360	0.154	0.361
REG2	0.232	0.422	7,973	0.236	0.425	0.237	0.425	0.234	0.423
REG3	0.413	0.499	7,973	0.415	0.492	0.415	0.492	0.416	0.493

Table 3
Robust Estimates of the Wage Model (OLS)

Variables	2000		2000-2006	
	Estimators	Standard Errors	Estimators	Standard Errors
COM	0.227**	0.015	0.292**	0.011
EDU1	-0.705**	0.043	-0.322**	0.039
EDU2	-0.571**	0.027	-0.170**	0.033
EDU3	-0.239**	0.029	0.089**	0.034
MST	0.068**	0.015	0.096**	0.012
GEN	-0.278**	0.014	-0.364**	0.011
HIST	-0.064**	0.020	-0.084**	0.017
BLK	-0.145**	0.018	-0.117**	0.014
UNI	0.175**	0.017	-0.038	0.019
URB	0.025	0.018	0.062**	0.015
SMSA	0.180**	0.032	0.195**	0.020
REG1	0.046	0.025	0.052*	0.021
REG2	-0.062**	0.023	-0.061**	0.019
REG3	-0.069**	0.022	-0.103**	0.018
Intercept	7.448**	0.046	7.084**	0.041

Note: * significant at 5% level, ** significant at 1% level.

Table 4:
Effect of Computer Skills on Wages after Controlling for Heterogeneity

Variables	Robust Fixed-Effects		Robust Random-Effect		Robust Instrument variable		Between Estimators	
	Estimators	Std. err.	Estimators	Std. err.	Estimators	Std. err.	Estimators	Std. err.
COM	0.045**	0.017	0.200**	0.011	1.217**	0.314	0.392**	0.024
EDU1	0.121	0.186	-0.365**	0.055	0.118	0.147	-0.325**	0.061
EDU2	0.139	0.163	-0.190**	0.047	0.043	0.073	-0.190**	0.029
EDU3	0.294*	0.141	0.096	0.049	0.118**	0.036	0.045	0.031
MST	0.009	0.023	0.087**	0.014	-0.105	0.067	0.084**	0.019
GEN			-0.360**	0.016	-0.409**	0.019	-0.372**	0.016
HIS			-0.084**	0.024	-0.021	0.030	-0.064**	0.024
BLK			-0.140**	0.019	0.018	0.046	-0.104**	0.021
UNI	-0.031	0.041	-0.025	0.025	-0.118**	0.025	-0.023	0.025
URB	0.057*	0.024	0.074**	0.016	0.047**	0.018	0.017	0.025
SMSA	0.005	0.024	0.114**	0.020	0.181**	0.024	0.030**	0.035
REG1	-0.057	0.093	0.037**	0.029	0.021	0.024	0.023	0.029
REG2	-0.082	0.071	-0.076**	0.026	-0.075**	0.022	-0.076**	0.027
REG3	-0.070	0.068	0.110**	0.024	-0.114**	0.022	-0.106**	0.025
Intercept	7.019**	0.149	7.228**	0.056	6.416**	0.242	6.964**	0.050

Note:The Hausman test rejects random-effects model over fixed-effects model $\chi^2_{(11)} = 221.16$ p = 0.000.
*significant at 5% level, **significant at 1% level.

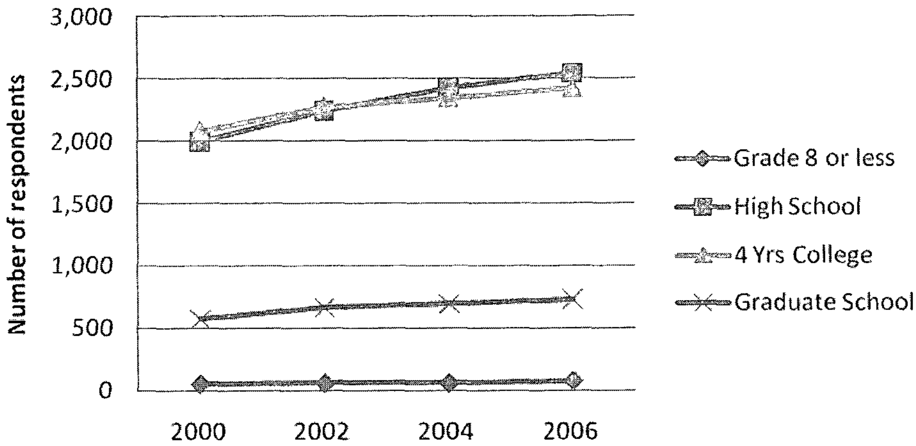

Figure 1
Home PC Ownership by Education

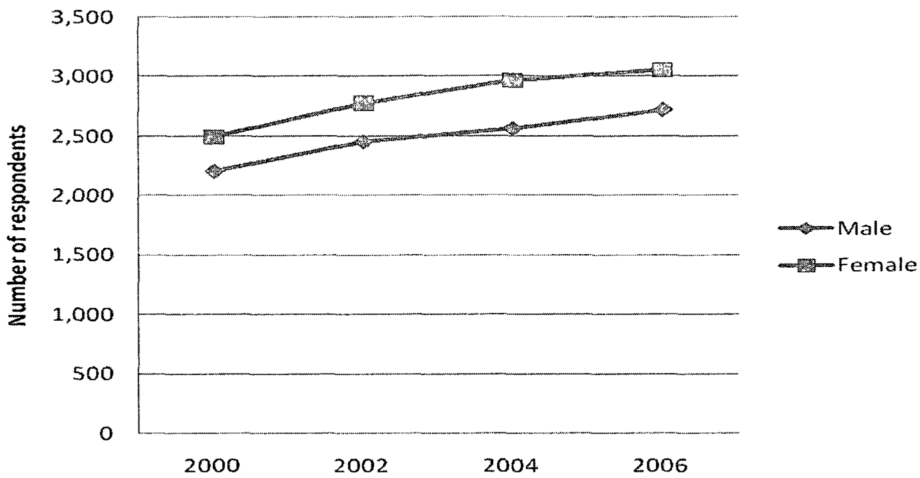

Figure 2
Home PC Ownership by Gender

CHAPTER 3

EFFECT OF EDUCATION ON HEALTH: A PANEL DATA STUDY FROM NLSY

Abstract

Using the NLSY79 panel data set, staggered every two years from 1994-2006 for a cross-section of 12,686 individuals, this paper investigates the effect of educational attainment on the health status of an individual as measured by the inability to work for health reasons. Based on the unique data set and the Arellano-Bond estimation methodology, it is found that educational attainment has a positive effect on the quality of an individual's health status. The present study also bridges the gap in the literature by using the robust fixed-effects model and Arellano-Bond to analyze the impact of education on health status.

Keywords: Education, Health Status, Fixed-Effects Model

JEL: I12, I20

Effect of Education on Health:
A Panel Data Study from NLSY

3.1. Introduction

The positive association between health and education is widely studied, reported, and accepted by health economists (Adams, 2002). The remaining dispute, however, relates to causation versus association (Berger and Leigh, 1989; Fuchs, 1982; Silles, 2009). One body of literature contends that education causes better health as it improves the technology of the health production function (Grossman, 1972). Education also increases the lifetime earnings of individuals, which makes the opportunity cost of becoming ill high and thus discourages them from participating in health-reducing activities (Cowell, 2006).

Another strand of the literature, however, casts doubt on whether better education leads to better health. This strand argues that the causation may be reverse or there may be no causal relationship between education and health, based on the assumption that there is a third missing factor such as the rate of time discount, heredity, or preferences that affect both education and health (Silles, 2009).

Furthermore, there are three main issues related to the education and health causal relationship. The first pertains to the definition of the heath status itself in that different variables are used as a measure of health status, including infant mortality rate, age-adjusted mortality rate, and life expectancy. The second has to do with the unobserved heterogeneity factor, which introduces bias in estimating the education parameter's effect on health. The third is the endogeneity problem associated with some of the determinants of the status including education and earnings. This essay re-examines these three issues using U.S. panel data from the National Longitudinal Survey of Youth 1979 (NLSY79) for the years 1994, 1996, 1998, 2000, 2004, and 2006 for a cross-section of 12,686 individuals in the survey. The data are observed in response to the NLSY79 survey question pertaining to the inability to work now due to illness. We use this variable as a measure of health status for the dependant variable as used by Berger and Leigh (1989). Most U.S. studies on the effect of education on health use cross-sectional data or one wave of longitudinal data and do not directly account for the fixed effects (see, for instance, Berger and Leigh, 1989; Adams, 2005). This essay uses a single-equation conventional model where education is the main explanatory variable instead of the two-equation model used by previous studies (Berger and Leigh, 1989; Arendt, 2005; Silles, 2009) for controlling the unobserved factors. We also test for heteroskedacity using the Breusch-Pagan approach and present robust standard errors. To the best of our knowledge, we have not found any published work that explicitly uses the robust fixed-effects model to control for unobserved heterogeneity and the Arellano-Bond (1991) approach to control for the endogeneity problem, which may confound the impact of education on health. The present study bridges the gap in the literature by using the fixed-effects model and Arellano-Bond model to analyze the impact of education on health status. We find no evidence that the causation runs from education to health in the fixed-effects model, while the Arellano-Bond model, which addresses both unobserved heterogeneity and the endogoneity problem, shows that education has a positive and significant impact on health, confirming the Grossman (1972) hypothesis.

The essay is organized as follows. The next section gives a review of selected literature. Section 3 discusses the data used for the study, while Section 4 presents the empirical model and results. The last section draws some conclusions based on the results.

3.2. Literature Review

Many studies have found the relation between the health of individuals and education (Silles, 2009) to be positive and statistically significant. Grossman (1972) articulated the idea that education improves the efficiency of the health production function, which in turn improves the health status. Cowell (2006) is of the opinion that education enhances the potential to earn and hence argues that individuals will avoid being involved in health-reducing activities, as the opportunity cost of being ill in the future is high. The causation here runs from education to health.

Some researchers are, however, skeptical of the fact that education causes health status to improve because there is a missing variable that affects both health and education. According to Fuchs (1982), this missing variable is the rate of discount, whereas Rosenzweig and Schultz (1983) argue that the missing variable may be an endowment such as a hereditary ability that affects both education and health. There may also be a case of reverse causation in the form of poor health that hinders attaining more education (Currie and Hyson, 1999). In his analysis, Grossman (1972) deals with the question of missing variables and reverse causation by using proxies such as parental education, test scores, and health at the high school level. However, his analysis does not deal with unobservable heterogeneity or endogeneity of education with health status (Arendt, 2005).

Wolfe and Behrman (1987) deal with endogeneity and unobserved heterogeneity by applying the within-family correlation technique. They collect data for sisters in Nicaragua since they are expected to have the same childhood background and control for unobserved elements related to their childhood background. They find that the mother's education has no significant effect on her children's health status. In another study, Behrman and Wolfe (1989), however, find that the women's education appears to make them healthier.

The unobservable heterogeneity and endogeneity problem is also typically dealt with using instrumental variables. Berger and Leigh (1989) use the per capita income and per capita expenditures in the state of birth as instruments. The result of their study shows that education significantly explains health status. The instruments may, however, be related to the expenditure on health, which might make them questionable (Arendt, 2005; Bound et al., 1995). Adams (2002) uses the quarter of birth as an instrument, since it affects one's educational attainment, and finds a positive but marginally significant effect of educational attainment on health. Lleras-Muney (2005) uses compulsory school and child labor laws in thirty states from 1915 to 1939 as instruments for education and finds that they have a significant effect in reducing mortality. Using panel data of school reforms as an instrumental variable for education in Denmark, Arendt (2005) finds the effect of education on the three alternative measures of health to be inconclusive. In a recent study, Silles (2009) also uses changes in compulsory schooling laws in the United Kingdom as an instrumental variable and finds a positive and significant effect of education on health.

As is evident from the above discussions, there is no definitive answer as to whether the instruments are weak or in some cases the results have very low precision. Previous studies of the relation between education and health are either based on a cross-sectional data framework (Berger and Leigh, 1989; Adam, 2002) or synthetic cohort analysis (Lleras-Muney, 2005). The Arendt (2005) study uses panel data but does not apply the fixed-effects model to control for unobserved heterogeneity. This study contributes to the literature by using the robust fixed-effects model and the Arellano-Bond (1991) model to address unobserved heterogeneity and enodogneity in analyzing the impact of education on health status.

3.3. Data

The empirical analysis of the present paper is based on data from the National Longitudinal Survey of the Youth 1979 (NLSY79), which is a nationally representative sample of 12,686 young men and women who were 14-22 years of age when they were first surveyed in 1979. These individuals were interviewed annually from 1979 through 1994 and have been interviewed on a biannual basis since 1994 (i.e., 1996, 1998, 2000, 2002, 2004, and 2006).

Although the original survey of the NLSY79 data set started with 12,686 respondents, the number of respondents decreased to 7,764 by the end of 2006 due to attrition. The data we used consider these observations as missing information, so they do not create a systematic bias. Unanswered questions are also considered missing data. In the 1994 survey, the respondents' age ranged from 29 to 37. The 1994 survey and those conducted afterwards ask the question *"Would your health limit the kind of work you do now?"* We use this variable as a measure of health status (dependent variable). The wording and structure of the questions regarding the variables used in this study remain the same and are exactly comparable.

Table 5 gives the definition of variables, while Table 6 presents the descriptive statistics of the variables used in the study. It is important to note that only 4 percent of the survey population consider that their health limits the kind of work they do now in 1994, which decreases to 2 percent in the year 2006. The family size also decreases from an average of 3.2 in 1994 to 2.9 in 2006. The educational attainment increases from 12 years in 1994 to 13 years in 2006. The empirical model and results are discussed in the next section.

3.4. The Empirical Model and Results

Our basic equation for determining the impact of education on health status is based on the standard formulation in most of the previous studies (Berger and Leigh, 1989; Sillies, 2009) as given below.

$$HST_{it}= \alpha_i + \gamma\, EDU_{it} + X_{it}\, \beta + \varepsilon_{it} \qquad (1)$$

where α_i is unobserved heterogeneity (also known as individual-specific effects). *HSTit* is a measure of the health status of individual *i* at time *t*, set as 1 when the individual is unable to work now for health reasons, zero otherwise. EDU_{it} is the educational attainment of individual *i* at time *t* measured as years of schooling completed. X_{it} denotes a vector of the control variables such as wages, gender, race, marital status, family size, residence in metropolitan area, and region of residence of individual *i* at time *t*. ε_{it} is the disturbance term, which accounts for omitted factors and other random errors.

Generally, when the dependent variable is a binary variable, which assumes a value of 1 or 0, nonlinear models such as the logit or probit models are preferable. In the present study, the robust logit estimation is conducted in addition to the linear probability model. In the linear probability model, education is found to be a statistically significant determinant of health status (Table 7). The education coefficient of – 0.002 implies that each additional year of education reduces the inability to work presently due to illness by only 0.20 percent. The robust logit model confirms that education is a statistically significant and positively related to the measure of the health.

Unobserved heterogeneity may be fixed or random over time. Here, time-related random unobserved heterogeneity or individual-specific effect heterogeneity is controlled by the robust random-effects model and between-effects estimator process. The individual fixed unobserved heterogeneity is controlled by the robust fixed-effects model. We use the robust fixed-effects Generalized Least Squares Method (FEGLS) for estimation. The results from the robust fixed-effects, random-effects, and between-effects models are reported in Table 8, which also reports the estimation of the logit model with fixed effects. Education is positively related to the health status variable. In the robust random-effects and between-effects estimators, it is a significant variable. In the robust fixed-effects model and logit model, education is found to be statistically insignificant.

The Hausman-test rejects the robust random-effects model in favor of the fixed-effects model. The robust random-effects and between-effects estimators show significant and larger coefficients because they are not taking into account the effects of the omitted individual-specific variables. Among notable control variables in the fixed-effects model, family size and residing in a metropolitan area or region 2 are found to have a significant effect on health status.

As noted above, the impact of education on health is controversial. The study is further extended to address the issue that health status explaining factors are either predetermined, endogenous, or both, and current-period heath status depends on its value in the past, a dynamic variant of equation (1) above, known as the Arellano-Bond (Fayissa et al., 2008) model, specified as follows:

$$\Delta HST_{it} = \delta\, \Delta HST_{it-1} + \gamma\, \Delta EDU_{it-1} + \Delta\, X_{it}\,\beta + \alpha_i + \varepsilon_{it} \qquad (2)$$

where ΔHST_{it} is the first difference of the health status of individual i during perod j; ΔHST_{it-1} is the lagged difference of dependent variables, ΔEDU_{it-1} is the lagged level and predetermined endogeneous variable, and ΔX_{it} is vector of exogenous variables. α_i and ε_{it} are assumed to be independent over all time periods for individual i. The term α_i represents individual-specific effects that are distributed independently and identically over the individuals, and ε_{it} is the noise stochastic distribution term and is also assumed to be distributed independently.

The coefficients of equation estimated by using the Arellano-Bond (1991) GMM estimator are reported in Table 5. The results show that when the unobserved heterogeneity and endogeneity are controlled, the education variable has a significant impact on the health status variable, reflecting that educational attainment has a positive effect on health. In other words, education improves the health of an individual. After controlling for endogeneity, we also find that increases in wages have a positive and significant effect on health status as reported in Table 9. All of the demographic variables used are found to have no significant effect on health status.

3.5. Conclusion

This essay has examined the question of causation and association between education and the health status of individuals. Previous studies of the relation between education and health status using cross-sectional data have been the subject of criticism for using weak instruments or not showing strong results (Arendt, 2005; Bound et al., 1995). This paper employs the fixed-effects model to control for the unobserved heterogeneity factor using NLSY79 panel data for 1994, 1996, 1998, 2000, 2004 and 2006 for a cross-section of 7,764 individuals[1] as well as the Arellano-Bond (1992) dynamic model to control for the endogeneity problem associated with the education and wages variables.

The fixed-effects results of this study show that educational attainment does not cause health status to improve, unlike some previous studies that suggested otherwise based on models that used instrumental variables (Berger and Leigh, 1989; Adams, 2002). Other interesting results are that family size and residence in metropolitan area are important factors that determine the health status of individuals. However, the fixed-effects model controls only for the unobserved heterogeneity factor and assumes that there may be no endogeneity of education or wages.

To address issues of both unobserved heterogeneity and endogeneity, we use the Arellano-Bond model, which is designed for few periods and a large number of individuals

[1] The original 1979 sample was 12, 686 individuals, but we ended up with 7,764 because of attrition and incomplete answers to survey questions.

in each panel. Based on the results of this methodology, it is found that education, wages and the lagged value of health status are significant determinants of health status, although the magnitude of the coefficients is very small. The present study uses a variety of models to investigate the impact of education, wages, and other control variables on health status. Both the OLS and logit models show that education and wages have a positive and significant impact on health as shown in Table 7. To control for unobservable individual heterogeneity and the time effect, we use the fixed-effects and random-effects models. The Hausman test that there is no difference between the estimates of the fixed-effects and random-effects models rejects the random-effects model in favor of the fixed-effects model, on which we base our findings as reported in Table 8. The results of the fixed-effects model show that education has a positive but not significant effect on health. This may be due to the endogeneity problem associated with the education variable. To correct for endogeneity, we use the Arellano-Bond dynamic model, which reveals that education has a positive and significant effect on health. Based on the Arellano-Bond (1991) analysis, it can be concluded that the Grossman (1972) interpretation of education as improving the efficiency of health production and the Cowell (2006) argument that the opportunity cost of future illness is higher for educated people, thus forcing them to refrain from health-harming activities, is confirmed. In other words, we find in this analysis evidence that education improves the health status of an individual after controlling for the individual's unobserved heterogeneity and the endogeneity problem associated with the education variable.

Table 5
Definitions of Variables

Variables	Variables Definition
HST	Inability to work due to health now (Health Status)
EDU	Years of education attainment
WAG	Wages $*10^{-4}$
FSZ	Family size of the individual
URB	Individual lives in urban area
MST	Marital Status
BLK	Black
HSP	Hispanic
GEN	Gender
SMSA	Standard Metropolitan Statistical Area
REG1	Region 1 (Northeast)
REG2	Region 2 (North Central)
REG3	Region 3 (South)

Table 6
Descriptive Statistics

	1994		1996		1998		2000		2002		2004		2006	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
HST	0.040	0.19	.046	.21	0.056	0.23	0.018	0.13	0.019	0.13	0.016	0.12	0.020	0.14
EDU	12.09	2.87	13.10	2.43	13.18	2.42	13.188	2.43	13.29	2.461	13.353	2.514	13.42	0.24
WAG	1224	1000	1371	1302	1493	1619	1685	2142	1806	1945	1862	1960	2051	3199
FSZ	3.160	1.54	3.21	1.55	3.27	1.5	3.21	1.55	3.179	1.527	3.069	1.504	2.96	1.47
URB	0.80	0.40	0.79	0.40	0.69	0.46	0.752	0.47	0.790	0.472	0.782	0.473	0.77	0.51
MST	0.57	0.50	0.57	0.49	0.59	0.49	0.589	0.49	0.597	0.490	0.597	0.490	0.59	0.49
BLK	0.19	0.39	0.18	0.38	0.18	0.38	0.187	0.390	0.182	0.386.	0.185	0.388	0.18	0.45
HSP	0.29	0.45	0.29	0.45	0.29	0.45	0.296	0.456	0.299	0.457	0.295	0.456	0.29	0.49
GEN	0.52	0.50	0.52	0.50	0.51	0.49	0.572	0.499	0.509	0.499	0.504	0.500	0.50	0.49
SMSA	0.29	0.45	0.14	0.34	0.30	0.46	0.285	0.451	0.272	0.445	0.298	0.457	0.33	0.47
REG1	0.13	0.34	0.15	0.36	0.15	0.35	0.152	0.359	0.151	0.358	0.153	0.360	0.15	0.36
REG2	0.24	0.43	0.23	0.42	0.23	0.42	0.239	0.426	0.243	0.429	0.244	0.429	0.24	0.43
REG3	0.41	049	0.47	0.49	0.41	0.49	0.411	0.492	0.414	0.492	0.413	0.492	0.40	0.49

Table 7
Robust OLS and Logit Model

	OLS Model		Logit Model	
Variables	Coefficients	Standard errors	Coefficients	Standard errors
EDU	-0.002**	0.0002	-0.122**	0.014
WAG	-0.018**	0.0038	-7.73**	0.811
FSZ	-0.0001	0.0005	-0.035	0.025
URB	-0.003**	0.001	-0.187*	0.081
MST	-0.010**	0.0016	-0.385**	0.081
BLK	-0.002	0.0019	-0.134	0.103
HIS	0.0015	0.0016	0.013	0.084
GEN	-0.005**	0.0013	-0.160*	0.069
SMSA	-0.0045**	0.0016	0.235**	0.078
REG1	-0.0047*	0.0022	-0.216	0.119
REG2	-0.0071**	0.0021	-0.437**	0.109
REG3	-0.0043*	0.0019	-0.321**	0.095
Intercept	0.077**	0.0048	-0.676**	0.228

*significant at 5%, **significant at 1%

Table 8
Effect on Health Status Controlling for Heterogeneity

	Logit Model	Linear Probability Model		
Variables	Fixed Effect	Robust Fixed-Effect	Robust Random Effect	Between Effect
EDU	-0.051(0.087)	-0.001 (0.001)	-0.004 (0.0005)**	-0.004(0.0005)8*
WAG	-0.887 (0.673)	-0.004(0.002)	-0.010 (0.002)**	-0.057(0.010)**
FSZ	-0.114(0.042)**	-0.002(0.001)**	-0.001 (0.0007)	0.002(0.001)*
URB	-0.179(0.118)	-0.002(0.002)	-0.003(0.001)*	-0007(0.004)
MST	0.003 (0.147)	0.0003 (0.002)	-0.006(0.002)**	-0.024(0.003)**
SMSA	0.427(0.133)**	0.007(0.002)**	0.006 (0.002)**	0.004(0.004)
REG1	-0.334 (0.619)	0.007(0.011)	-0.005(0.004)	-0.003(0.004)
REG2	-0.885(0.446)*	-0.022(0.011)	-0.008(0.004)*	-0.006(0.004)
REG3	-0.298 (0.387)	-0.009 (0.009)	-0.005 (0.003)	-0.004(0.003)
Intercept		0.052(0.025)**	0.100(0.009)**	0.112(0.009)**

Note: In parentheses are robust standard errors.
Hausman test: rejects random-effects model in favor of fixed-effects model
$\chi^2_{(9)} = 69.08$, p = 0.000
*significant at 5%, **significant at 1%

Table 9
Arellano-Bond Dynamic Panel Data Estimation Results

Variables	Coefficient Estimates	Standard Errors
HST(LD)	-1.29e^(-14)**	1.61e^(-15)
HST(L2D)	-7.12e^(-15)**	1.52e^(-15)
EDU(D(1))	-9.23e^(-16)**	2.61e^(-16)
WAG(D(1))	-1.16e^(-21)**	6.89e^(-21)
FSZ(D(1))	-1.99e^(-18)	1.86e^(-17)
URB(D(1))	1.12e^(-18)	3.78e^(-17)
MST(D(1))	1.10e^(-18)	6.17e^(-17)
SMSA(D(1))	4.95e^(-17)	5.43e^(-17)
REG(D(1))	-1.04e^(-17)	2.63e^(-16)
REG(D(2))	-1.01e^(-16)	2.32e^(-16)
REG(D(3))	-1.04e^(-17)	1.98e^(-16)

Sargan test of over-identifying restrictions: $\chi^2_{(8)} = 13{,}786$, p> $\chi^2_{(8)} = 0.00$
Arellano-Bond test of the null of AR(1) residual errors z = -18.63**
Arellano-Bond test of the null of AR(2) residual errors z = -3.71**
**Significant at 1%. While the suffix *D(1)* after each variable denotes the number of time each variable was differenced, *LD* denotes the lagged difference. The *HST* is treated aspredetermined, while *EDU* is treated as an endogenous variable.

CHAPTER 4

IMPACT OF EDUCATION ON LIFESTYLES: A PANEL DATA STUDY FROM NLSY79

Abstract

This essay investigates the effect of education on different lifestyle variables using NLSY79 panels for 1992, 1994, and 1998. The lifestyle variables are smoking, drinking, marijuana use, and cocaine use. The analysis addresses the joint determination of lifestyle variables within the framework of the Seemingly Unrelated Regression (SUR) model. Most previous studies use the instrumental variable technique for controlling individual unobserved heterogeneity. In this essay, unobserved heterogeneity is controlled by the robust fixed-effect model extended to SUR. It is found that educational attainment has no significant effect on the lifestyle choices of individuals. The present study bridges a gap in the literature by using a robust fixed-effects model within the SUR framework to analyze the impact of education on lifestyle choices.

Keywords: Education, Smoking, Drinking, Marijuana Cocaine, Fixed-Effects Model, SUR Model

JEL Classification: I1, I2, I10, I12, C30

Impact of Education on Lifestyles: A Panel Data Study from NLSY79

4.1. Introduction

There have been long quests by both epidemiologists and economists to explain inequalities in health. Researchers appear to reach the general consensus that while access to medical care is important, it does not fully explain inequalities in the health status of individuals (Lleras-Muney, 2005; Folland et al., 2001). Economists have widely studied and recognized that there is a positive relation between health and education (Grossman, 1972; Ross and Wu, 1995). They have also found evidence that education is strongly correlated with most health-related behaviors or lifestyle factors such as smoking, drinking, and substance abuse. These lifestyle factors are assumed to be inputs in the individual's production function of health (Park and Kang, 2008). Disputes that persist, however, relate to the issue of causation versus association. One body of literature contends that causation runs from education to health or health-related behavior. Grossman (1972) articulates the idea that education improves the efficiency of health production. Similarly, Cowell (2006) argues that educated people will choose to have improved health-related behaviors or lifestyles, as the opportunity cost of being ill in the future is high.

Another body of literature, however, casts doubt on whether better education necessarily leads to better health or lifestyle choices that improve health. One justification of this view rests on the notion that the causation may be reverse. In other words, is it better health that leads to better schooling (Currie and Hyson, 1999), or do lifestyle choices such as substance abuse affect educational attainment (Register and Grimes, 2001). Still other studies argue that there may be no causal relationship since there may be other factors such as the rate of discount, heredity, or preferences affecting both education and health or health-related behavior (Fuchs, 1982; Rosenzweig and Shultz, 1983; Farrell and Fuchs, 1982). Fuchs (1974) is also of the opinion that personal lifestyle is the most important factor causing great variations in health.

The present paper investigates the impact of education on the lifestyle variables that are assumed to be determinants of health by considering three main issues: (1) the definition of lifestyle, (2) the unobserved heterogeneity among agents or individuals that makes the empirical results biased, and (3) the joint determination of lifestyle variables treated as choice variables for individuals. Previous studies have used either only one choice variable such as smoking (deWalque, 2007; Sander, 1995) or multi-lifestyle variables such as smoking; drinking; substance use; having regular breakfast, exercise, and medical checkups; or restful patterns of sleeping in finding the impact of education on lifestyle choices (Kenkel, 1991, 1997; Park and Kang, 2008). For the purpose of this essay, only smoking, marijuana use, cocaine use, and drinking are used as proxies for lifestyle variables.

In earlier studies, the unobserved heterogeneity of education among individuals or agents was not considered (Kenkel, 1991). Recent studies have, however, controlled the unobserved heterogeneity among agents and the endogeneity of the education variable by using the instrumental variable technique (IV). The use of the instrumental variable such as family background was criticized (Grossman, 2005). The trend in the choice of the instrumental variables has evolved to the wide use of institutional characteristics (Park and Kang, 2008). The difficulties associated with finding the right instruments are, however, well documented in the literature (Bound et al., 1995). This essay takes a different approach to avoid the controversy surrounding the selection of instruments by adopting the fixed-effects method on panel data drawn from the NLSY79.

In dealing with the multi-lifestyle variables listed above, Kenkel (1991) estimates each equation separately without considering their joint determination, in contrast to Park and Kang (2008). This essay also addresses the issue of the joint determination of the lifestyle variables within the empirical framework of the Seemingly Unrelated Regression (SUR) model. It is shown that estimating each equation using the fixed-effects model is appropriate if the explanatory variables or regressors are identical in each equation. This essay utilizes the fundamentals of the SUR model, which is generally underutilized in educational research, as argued by Beasley (2008) and Green (2008: 267), and extends the fixed-effects model into the SUR.

In the literature, the relation of the random- and fixed-effects models to SUR models are discussed by Avery (1977), Wooldridge (2002: 272-274), and Blackwell (2005), but the elimination of fixed effects from SUR regressions is not discussed explicitly. This essay is different from previous studies in two respects: (1) it incorporates the fixed-effects model into the SUR framework, and (2) it uses the robust fixed-effects technique to address the issue of individual unobserved heterogeneity arising from the cross-correlation and within-correlation of lifestyles.

The essay uses panel data from the National Longitudinal Survey of Youth 1979 (NLSY79). Only two-period panels (1992 and 1994) are used for determining the impact of education on smoking, alcohol consumption, marijuana, and cocaine due to data unavailability for the drinking variable for 1998. We, however, use the 1992, 1994, and 1998 panel data to examine the impact of education on the other three variables (smoking, marijuana use, and cocaine use). Overall, the study shows that education does not have a significant effect on lifestyle choices after controlling for individual unobserved heterogeneity.

The essay is organized as follows. The next section gives a brief review of the literature. Section 3 describes the data and methodology. The empirical results are presented in section 4. The final section draws conclusion based on the results.

4.2. Review of Literature

Health care is one of the factors in maintaining health. Lifestyle is also considered to be an important determinant of health (Folland et al., 2001). Fuchs (1979) compared Utah and Nevada in terms of death rates and found that such factors as abstinence from the use of tobacco and alcohol will increase indivituals' longevity. In another study, Fuchs (1982) also argued that personal lifestyle is a significant determinant of health, while the Rosenzweig and Schultz (1983) study shows that maternal cigarette smoking has a significant negative effect on newborn birth weight. Joyce et al. (1992) have also found that illicit substance abuse by pregnant women results in significant harm to the newborn.

Many studies have addressed the effect of education on one lifestyle variable such as smoking. Among them, Sander (1995) finds that education has a negative effect on smoking after accounting for the endogeneity of the education variable using the instrumental variable (IV) technique. Since the instruments used are related to the family-background variables, however, they were criticized as being weak instruments (Grossman, 2005). Examining the effect of mother's education on birth outcome and smoking, using the availability of college in maternal county as an instrumental variable for the mother's education, Currie and Moretti (2003) find a positive effect of education on birth weight and a negative effect of education on smoking during pregnancy. After controlling for endogeneity using Vietnam veteran's status as an instrumental variable, Grimard and Parent (2007) find no significant effect of education on smoking, while de Walque (2008) finds that education discourages smoking by using the risk of being drafted to the Vietnam War as an instrumental variable for men's college education.

Using cigarette smoking, alcohol consumption, and exercise as lifestyle variables, Kenkel (1991) finds it is health knowledge that affects lifestyle but draws no conclusion about schooling as a factor affecting the above lifestyle variables. On the other hand, Contoyannis and Jones (2004) use a multivariate probit model (of the Maximum Simulated Likelihood variety) for determining the impact of discrete lifestyle choices on self-assessed health. They find that lifestyle choices affect health and that education affects lifestyle choices. Park and Kang (2008) examine health-related multiple lifestyle choices such as regular checkups, exercise, smoking, and alcohol consumption. Using the instrumental variable technique, they find that education is not a significant determinant of smoking or drinking, but they find that education has a significant effect on regular exercise and checkups. As is evident from the above discussions, the effect of education on lifestyle choices and, indirectly, health status is mixed. This essay attempts to provide some evidence on the effect of education on lifestyle choices using the fixed-effects model to control for individual unobserved heterogeneity. In the present analysis, we extend the fixed-effects model to the SUR framework. More specifically, this approach takes into consideration the cross- and within-correlation problems that arise among lifestyle variables and individuals based on panel data from NLSY79, described in the following section.

4.3. Data and Methodology

The data set drawn from NLSY79 for the present study is a nationally representative sample of 12,686 of men and women who were 14-22 years old when they were first surveyed in 1979. These individuals were interviewed annually from 1979 through 1994 and biannually since 1996. The collection of data for drug use began in 1988 and then was repeated in 1992, 1994, and 1998. There were variations in some of the survey questions as well as the introduction of new variables with respect to drug use in the different surveys. For the purpose of the present analysis, however, data regarding the four lifestyle variables were extracted only from the 1992 and 1994 surveys. The data sets for the years 1992, 1994, and 1998 are used to analyze the three lifestyle variables (smoking, marijuana use, and cocaine use). The NLSY79 survey started with 12,686 respondents, but the respondents decreased to 8,794 by the end of 1994 and 8,403 by the end of 1998 due to attrition. The present essay treats those observations as missing information to avoid systematic bias. Unanswered questions are also considered as missing data.

The definitions of variables are given in Table 10. The dependent variables are in binary form. Among the explanatory variables, education and wage are in continuous form, while the rest of regressors are dummy variables. The descriptive statistics are presented in Table 11. The data suggest that the trend in smoking, drinking, marijuana, and cocaine use are declining as shown in Figures 1 through 4. The years of education have increased from 12.8 years to 13.0 years, while wages declined from $16.70 in 1992 to $14.76 in 1998 (a decline of about 12 percent). Almost 50 percent of the respondents were male, 25 percent were Black, and 15.7 percent were Hispanic. About 58 percent of the respondents were married, and more than 70 percent of the respondents lived in urban locations.

Figures 3 through 6 show that the trend for smoking, marijuana, and cocaine use has declined steeply for both genders. The drinking variable decreased for both genders but at a slow rate. The figures also reveal that males dominate in drinking, smoking, marijuana, and cocaine use. Whether this declining trend for these variables is due education is the objective of this study. As discussed above, this essay deals with the impact of education, wages, and other demographic variables (gender, marital status, race, and location of residence) on various lifestyle choices by heterogeneous individuals with important implications for their health outcomes. Since the effect of education and other demographic factors on multiple lifestyle variables such as smoking, drinking, marijuana use, and cocaine use involves multiple equations that on the surface appear to be independent of each other but in fact involve cross-correlations, we employ the SUR model as developed by Zellner (1962).

In the present analysis, there are two sets of equations. In the first set, there are four equations; in the other set, there are three equations. We initially set up the SUR model for the four dependent lifestyle variables and then estimate the fixed-effects model. The analysis can, however, be generalized to any number of structural equations that can be specified as follows:

$$\mathbf{Y_{it,j} = \beta_j X_{it} + €_{it,j}} \qquad \text{where } \mathbf{j=1, 2,3,4.} \qquad \mathbf{(1)}$$

where $\mathbf{Y_{it,j}}$ is a vector of the dependent lifestyle variable for individual i at time t, and $_j$ indexes the equation number. $\mathbf{X_{it,j}}$ denotes the control variables such as education, gender, race, marital status, wages, and location of residence. $€_{it,j}$ is the vector the disturbance terms. Model (1) can be illustrated in a vector matrix form as below:

$$\begin{bmatrix} Y_{it,1} \\ Y_{it,2} \\ Y_{it,3} \\ Y_{it,4} \end{bmatrix} = \begin{bmatrix} X_{it,1} & 0 & 0 & 0 \\ 0 & X_{it,2} & 0 & 0 \\ 0 & 0 & X_{it,3} & 0 \\ 0 & 0 & 0 & X_{it,4} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} + \begin{bmatrix} €_{it,1} \\ €_{it,2} \\ €_{it,3} \\ €_{it,4} \end{bmatrix} \quad (2)$$

(i)(t)(4)×1 (i)(t)(4) ×K K×1 (i)(t)(4)×1

where $K = \sum K_j$, and K_j is the number of explanatory variables in equation j. In the absence of individual-specific errors, and assuming strict homoscedasticity and that disturbance terms are correlated across but uncorrelated within equations, we can express the covariance matrix of the error terms as:

$$E(€_{it,j}\, €'_{it,j}) = \boldsymbol{\Omega} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} \end{bmatrix} \otimes \mathbf{I}_2$$

If the set of regressors are not identical and $\boldsymbol{\Omega}$ is not known, then the SUR model can be estimated by the Feasible Generalized Least Squares Method (FGLS); if $\boldsymbol{X}_{it,1} = \boldsymbol{X}_{it,2} = \boldsymbol{X}_{it,3} = \boldsymbol{X}_{it,4}$, however, then a System Ordinary Least Squares Method (SOLS) is equivalent to FGLS (Green, 2008, 257-258; Wooldridge, 2002, 148-150). SOLS is the ordinary least squares (OLS) estimation, equation by equation.

In the present study, the regressors in all equations are assumed to be identical. The OLS method of estimation, equation by equation, is applied, which is equivalent to the FGLS method (which considers the cross-correlation among equations). A test for heteroscedasticity reveals its presence, and we use the Breusch-Pagan approach to correct it. In this study, we use the robust OLS equation by equation estimates of the coefficients to compare our results to the robust fixed-effects model.

The individual-specific effects or fixed effects in the above-stated SUR model in equation (2) are described as follows. The disturbance terms $\boldsymbol{\epsilon}_{\mathbf{itj}}$ account for the individual-specific or omitted variables and other random errors. They are decomposed into individual specific errors $\mathbf{c}_{i,j}$ and random errors or "noise" term $\mathbf{u}_{\mathbf{it,j}}$. The vector matrix form becomes as follows:

$$\begin{bmatrix} Y_{it,1} \\ Y_{it,2} \\ Y_{it,3} \\ Y_{it,4} \end{bmatrix} = \begin{bmatrix} X_{it,1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & X_{it,2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & X_{it,3} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & X_{it,4} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} + \begin{bmatrix} c_{i,1} \\ c_{i,2} \\ c_{i,3} \\ c_{i,4} \end{bmatrix} + \begin{bmatrix} \mu_{it,1} \\ \mu_{it,2} \\ \mu_{it,3} \\ \mu_{it,4} \end{bmatrix} \quad (3)$$

(i)(t)(4)×1 (i)(t)(4) ×K K×1 (i)(t)(4)×1

$K = \sum K_j$, and K_j is the number of the explanatory variable in equation j.

As in Avery (1977), we consider $c_{i,j}$ in each equation as fixed effects instead of random effects and extend it to the multiple equations of the SUR model. While Blackwell (2005) estimates the fixed-effects coefficients $c_{i,j}$ in the SUR model, we initially remove the$c_{i,j}$ as done in the fixed-effects model (Green, 2008, 190). The across- and within-correlations are assumed to take the following form:

$$E(\mu_{it,j}\mu_{it,l}) = \boldsymbol{\sigma}_{jl}$$

$$E(c_{i,j}\, c_{i,l}) = \theta_{jl}$$

$$E(\mu_{it,j}\mu_{rs,l}) = \mathbf{0} \quad \mathbf{i \neq r}$$

$$E(\mu_{it,j}\mu_{rt,j}) = \mathbf{0} \quad \mathbf{i \neq r}$$

$$E(c_{r,j}\, c_{s,j}) = \emptyset_{rs}$$

There are two ways to remove $c_{i,j}$. One way is to take the first difference, while the other way is to use the fixed-effects transformation following Wooldridge (2002). For this essay, the estimation procedure is done in two steps. First, the fixed-effects transformation is done, which first averages each equation j over the period to get the cross-sectional effect for each equation j as given below,

$$\bar{Y}_{i,j} = \beta \bar{X}_{i,j} + c_{i,j} + \bar{\mu}$$

which is subtracted from each j[th] equation.

The system of equations now becomes

$$\begin{bmatrix} \Delta Y_{it,1} \\ \Delta Y_{it,2} \\ \Delta Y_{it,3} \\ \Delta Y_{it,4} \end{bmatrix} = \begin{bmatrix} \Delta X_{it,1} & 0 & 0 & 0 \\ 0 & \Delta X_{it,2} & 0 & 0 \\ 0 & 0 & \Delta X_{it,3} & 0 \\ 0 & 0 & 0 & \Delta X_{it,4} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} + \begin{bmatrix} \Delta \mu_{it,1} \\ \Delta \mu_{it,2} \\ \Delta \mu_{it,3} \\ \Delta \mu_{it,4} \end{bmatrix} \qquad (4)$$

where Δ reflects that the variable is in deviation form. In the second step, the FGLS is applied to the system of equations in (4). If the regressors are identical, then SOLS is appropriate and is equivalent to applying the fixed effects equation by equation in model (2) above. The extension of the fixed-effects model to the SUR model is a contribution to the literature of this essay to the best of the author's knowledge. To take care of heteroscedasticity, the robust fixed-effects model is applied to each equation. The results from the application of the above techniques are discussed in the next section.

4.4. Empirical Results

The results from the system of robust OLS regressions using the data set for years 1992 and 1994 to obtain the estimates of the four lifestyle variables and data set for years 1992, 1994 and 1998 are presented for the three lifestyle variables in Table 12. The results show that education discourages smoking, marijuana use, and cocaine use for both sets of data. For the drinking lifestyle, however, the educational attainment variable has a positive and statistically significant effect. Kenkel's (1991) study also shows that education has a negative impact on smoking and positive effect on drinking. Using South Korean data, Park and Kang (2008) also obtain similar results for the smoking and drinking variables when individual unobserved heterogeneity is not controlled. The result of the robust OLS estimate showing that education has a negative impact on smoking is in line with other works by Sander (1995), Leigh and Dhir (1997), and de Walque (2008) when they apply the OLS method.

The robust OLS results show that for each additional year, educational attainment reduces smoking, marijuana use, and cocaine use by 5 percent, 0.4 percent, and 0.1 percent, respectively. The analysis for the 1992 and 1994 periods indicates that education has a significantly positive effect on alcohol drinking, perhaps confirming recent medical information suggesting that moderate drinking of one or two servings of alcohol daily may promote the health status of individuals (Rimm et al., 1996; Davies et al., 2002; DHHS, 2005). The robust fixed-effects model is applied to estimate equation by equation, and the results are reported in Table 13. Wages are still insignificant in the robust fixed-effects model as in the OLS, while the impact of education on lifestyle choices becomes insignificant, contrary to the robust OLS results.

4.5. Conclusion

This essay has explored the effect of educational attainment on lifestyle choices by considering the correlation among lifestyle variables. The across-correlation among equations is discussed from the perspective of the SUR model. It has been noted above that if each equation has an identical set of explanatory variables than estimation of the SOLS method is equivalent to the estimation by FGLS for the SUR model. It is found that there is a negative and significant effect of educational attainment on smoking, marijuana use, and cocaine use and a positive effect on drinking when unobserved heterogeneity is not controlled.

The essay addresses the issue of heterogeneity by extending the robust fixed-effects model to multiple equations within the SUR framework. When the heterogeneity of individuals is controlled, the results are very different from the analysis based on the robust OLS method. It is found that education has no statistically significant effect on lifestyle variables, although it has a negative overall impact. The negative and significant effect for the educational attainment variable in the robust OLS analysis may be due to the unobserved heterogeneity factor. The time trend shows that, over the period of the study, the lifestyle variables have negative trends. After controlling for heterogeneity and heteroskedacity using the robust fixed-effects model, it can be concluded that factors other than education may have been responsible for the relationship. As Kenkel (1991) argued and concluded, it is health knowledge that affects lifestyle variables; also, there is a possibility that legal policy and its strict implementation may be the cause in explaining the negative trend of lifestyles variables used in the present analysis.

Finally, it is evident from the present analysis that education is not a major factor in influencing the choice of healthy lifestyles. If lifestyles are important inputs in the health production function, then education is insignificant in affecting health through lifestyle choices. In the present analysis, we do not find evidence that education improves the efficiency of a healthy lifestyle.

Table 10
Definitions of Variables

Variables	Variables type	Variable Definition
SMK	*Dummy*	Is smoking
MRJ	*Dummy*	Is using marijuana
COC	*Dummy*	Is using cocaine
DRK	*Dummy*	Is drinking alcohol
EDU	*Continuous*	Number of year of school attainment
WAG	*Continuous*	Hourly payment*10^{-4}
GEN	*Dummy*	Is Female
HIS	*Dummy*	Is Hispanic
BLK	*Dummy*	Is Black
MST	*Dummy*	Marital Status is "Married"
URB	*Dummy*	Lives in Urban Area
SMA	*Dummy*	Lives in Standard Metropolitan Statistical Area

Table 11
Descriptive Statistics

Variables	1992 Mean	1992 Standard deviation	1994 Mean	1994 Standard deviation	1998 Mean	1998 Standard deviation
SMK	0.324	0.468	0.326	0.469	0.306	0.461
MRJ	0.067	0.251	0.083	0.277	0.055	0.228
COC	0.011	0.106	0.013	0.115	0.008	0.090
DRK	0.630	0.482	0.596	0.490	-	-
EDU	12.848	2.451	12.910	2.451	13.046	2.462
WAG	16.706	39.492	12.259	9.941	14.765	15.983
GEN	0.495	0.499	0.495	0.499	0.495	0.499
HIS	0.157	0.364	0.157	0.364	0.157	0.364
BLK	0.250	0.433	0.250	0.433	0.250	0.433
MST	0.463	0.498	0.250	0.433	0.265	0.441
URB	0.816	0.387	0.805	0.395	0.700	0.457
SMA	0.778	0.415	0.812	0.390	0.800	0.399

Table 12
Robust Estimators Without Controlling Heterogeneity

	Dependent Variables							
	SMK		*MRJ*		*COC*		*DRK*	
Variables	1992-1994	1992-1998	1992-1994	1992-1998	1992-1994	1992-1998	1992-1994	1992-1998
EDU	-0.050**	-0.048**	-0.004**	-0.004**	-0.001**	-0.001**	0.009**	
	(0.001)	(0.001)	(0.0009)	(0.0007)	(0.0003)	(0.0002)	(0.001)	
WAG	-0.001**	-0.001**	-0.0004**	-0.0004**	-0.00005**	-0.0006*	0.0008**	
	(0.0002)	(0.0003)	(0.0001)	(0.0001)	(0.00001)	(0.00002)	(0.0001)	
GEN	-0.031**	-0.029**	-0.043**	-0.041**	-0.007**	-0.006**	-0.171**	
	(0.008)	(0.006)	(0.004)	(0.003)	(0.001)	(0.001)	(0.008)	
HIS	-0.0147**	-0.149**	-0.0.34**	-0.033**	0.001	0.001	-0.057**	
	(0.010)	(0.008)	(0.006)	(0.004)	(0.002)	(0.002)	(0.011)	
BLK	-0.072**	-0.070**	-0.037**	-0.036**	-0.002	-0.002	-0.142**	
	(0.010)	(0.008)	(0.005)	(0.004)	(0.002)	(0.001)	(0.010)	
MST	0.118**	0.119**	0.049**	0.050**	0.011**	0.011**	0.084**	
	(0.008)	(0.007)	(0.005)	(0.004)	(0.002)	(0.001)	(0.009)	
URB	-0.007	0.014	-0.001	0.007	0.005	0.002	0.064**	
	(0.013)	(0.009)	(0.007)	(0.004)	(0.002)	(0.001)	(0.014)	
SMA	0.028*	0.082	0.019**	0.013**	-0.0007	0.001	0.028	
	(0.012)	(0.009)	(0.006)	(0.004)	(0.002)	(0.001)	(0.013)	
Intercept	0.968**	0.944**	0.133**	0.127**	0.024**	0.020**	0.536**	
	(0.024)	(0.0.20)	(0.013)	(0.010)	(0.005)	(0.003)	(0.026)	

Note: Figures in parentheses are robust standard errors. * Significance of 5%, ** significance of 1%.

Table 13
Robust Fixed-Effects Estimators

	Dependent Variables							
	SMK		*MRJ*		*COC*		*DRK*	
Variables	1992-1994	1992-1998	1992-1994	1992-1998	1992-1994	1992-1998	1992-1994	1992-1998
EDU	-0.0008	-0.002	-0.014	-0.005	-0.002	0.002	-0.021	
	(0.0147)	(0.007)	(0.012)	(0.004)	(0.005)	(0.002)	(0.020)	
WAG	-0.00008	-0.00006	-0.00006	-0.00004	-0.000008	-0.000005	0.0002	
	(0.00009)	(0.00004)	(0.00009)	(0.00005)	(0.00003)	(0.00006)	(0.0002)	
MST	0.0163	0.021*	0.018	0.020**	0.004	0.003	0.050*	
	(0.0144)	(0.009)	(0.014)	(0.007)	(0.006)	(0.003)	(0.022)	
URB	0.0257	0.016*	-0.034	0.004	0.002	-0.006	0.038	
	(0.0297)	(0.008)	(0.028)	(0.006)	(0.010)	(0.002)	(0.038)	
SMA	-0.0002	-0.001	0.013	0.009	-0.005	-0.001	-0.0008	
	(0.019)	(0.014)	(0.014)	(0.010)	(0.006)	(0.004)	(0.030)	
Intercept	0.297	0.320**	0.263	0.123	0.044	-0.020	0.865**	
	(0.193)	(0.093)	(0.166)	(1.86)	(0.067)	(0.037)	(0.271)	

Note: Figures in parentheses are robust standard errors. * Significance of 5%, ** significance of 1%.

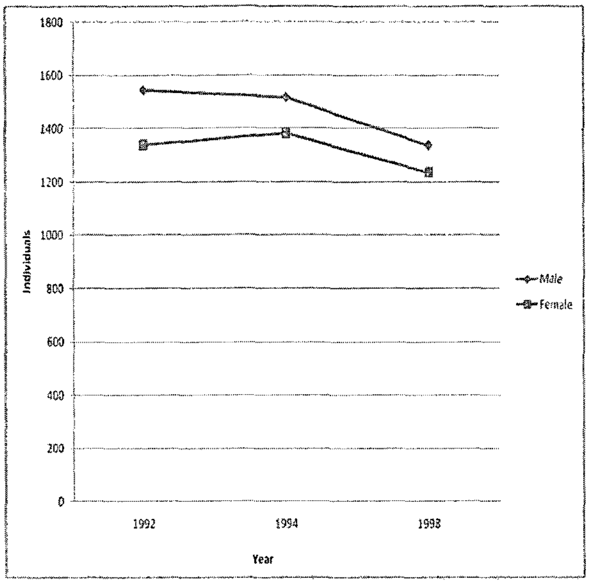

Figure 3
Smoking Trend

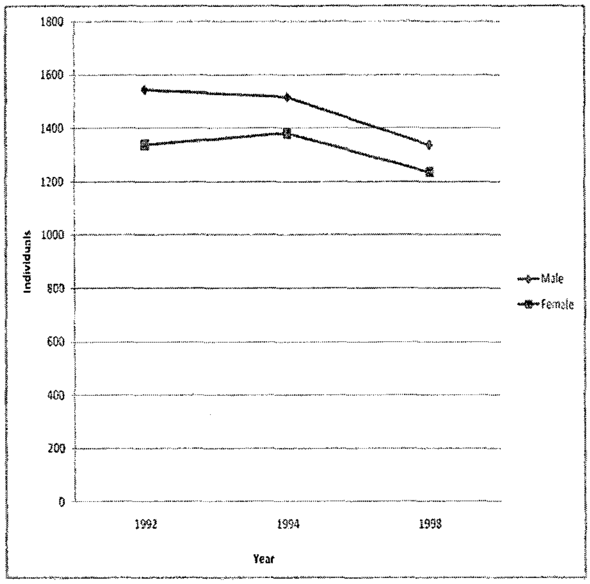

Figure 4
Marijuana Use Trend

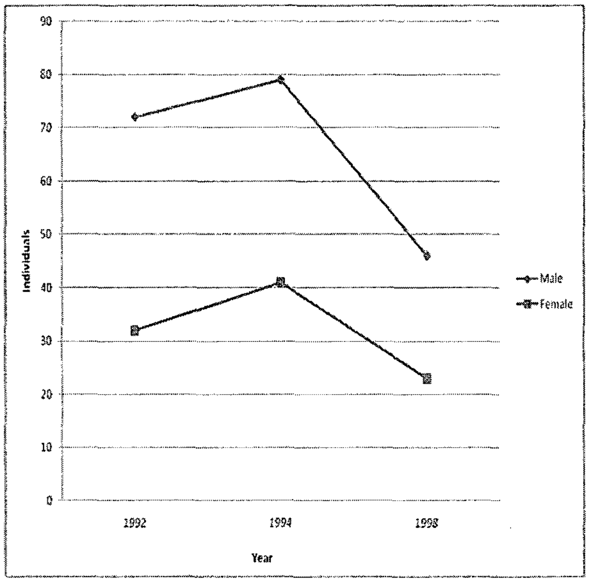

Figure 5
Cocaine Use Trend

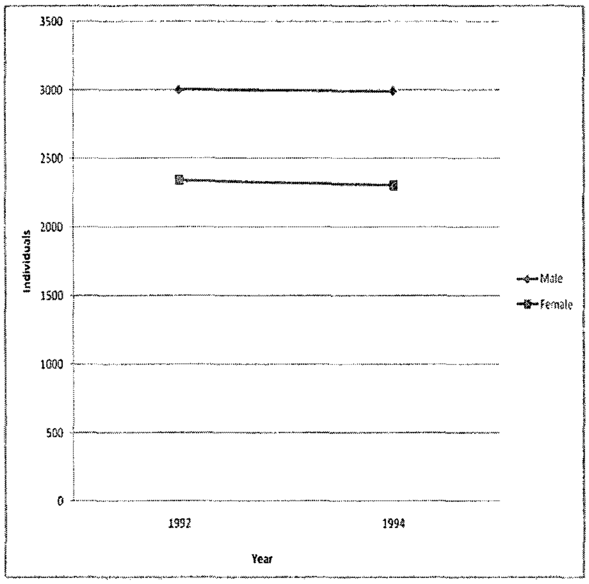

Figure 6
Drinking Trend

CHAPTER 5

CONCLUSION

Essay one in chapter two investigates the "Impact of Computer Skills on Wages" in the U.S. based on panel data drawn from the National Longitudinal Survey of Youth 1979(NLSY79). A wide variety of studies that had been done for the U.S. suggest that there is a pecuniary premium on computer skills. Implicit in these studies is that the single most important factor bringing a great change in technology is computer skills and that such technological change is easily adapted by highly educated individuals who are rewarded by higher wages for their skills in information technology. The wage gap that exists between those individuals who possess advanced computer skills and those who are less skilled is known as the skill-biased technological hypothesis (SBTC). Defining computer skills as having a computer with Microsoft Windows or NT at home and using the fixed-effects model and instrumental variable technique, the study finds that individuals possessing computer skills indeed earn a wage premium, thus confirming the SBTC hypothesis.

The second essay, in chapter 3, focuses on another aspect of investment in capital with respect to the effect of education on health using panel data drawn from the NLSY. The essay uses the inability to work due to health reasons as a proxy for health status. Controlling for unobserved individual heterogeneity by the robust fixed-effects model, the study finds that the impact of education on health status is insignificant. However, after controlling for the endogeneity problem arising from the interaction between educational attainment and the proxy for health status by employing the Arellano-Bond dynamic model, the study finds that educational attainment has a positive and statistically significant effect on the measure of health status. The study bridges the gap in the literature by using the robust fixed-effects model and the Arellano-Bond model to analyze the impact of education on health status after controlling for unobserved individual heterogeneity and the endogeneity problem arising from the interaction between education and the measure of health status. Future studies can explore the impact of educational attainment on health by constructing new proxies for health status such as blood pressure, heart problems, etc., as more survey data become available or as alternative estimation methodologies are constructed to improve the results of this study. In addition, the impact of such variables as health knowledge and public health policies can be included as determents of health in further research work.

Essay three explores the effect of educational attainment on different lifestyle variables using NLSY79 panels for 1992, 1994, and 1998 in chapter 4. Using lifestyle variables such as smoking, drinking, marijuana use, and cocaine use, the study addresses the joint determination of lifestyle variables within the framework of the Seemingly Unrelated Regression (SUR) model. After controlling for unobserved individuals heterogeneity by the robust fixed-effects model extended to the SUR model, the study finds that educational attainment does not necessarily have a significant effect on lifestyle choices. While future study with an adequate database and alternative methodologies may find different results and explanations, the finding of this essay suggests that it is health knowledge that affects lifestyle choices (such as warning labels on cigarette and alcohol packaging and nutritional information on processed food labels) rather than the educational attainment of individuals. The marginal contribution of this essay to the literature is the use of the robust fixed-effect model in the context of the SUR model to analyze the impact of cross- and within-correlations of educational attainment on lifestyle choices.

All three essays have practical implications. The first essay provides direction to policymakers as to what type of educational attainment or skills are most important to enhance the earning of individuals and also to ease the scarcity constraint of skills in the production process of firms using new technology, resulting in more output for the economy. The second essay tries to determine whether education or some other variables are important in explaining health for individuals so that policy emphasis should be placed on those factors that are important in determining health. The third essay suggests that other factors such as government policies and laws or health knowledge rather than education may be important factors in curbing or discouraging individuals from adopting lifestyles injurious to their health.

References:

Adams, S. (2002). Education attainment and health: Evidence from a sample of older adults, *Education Economics*, 10, 97-109.

AFP (November 5, 2009). Windows 7 debut a hit: NPD. Retrieved on April 6, 2010. from http://www.google.com/hostednews/afp/article/ALeqM5ixBXuAiKbUyygmDOdf5owPAjWjSg.

Arellano, M., and Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *Review of Economic Studies*, 58, 277-297.

Arendt, J.N. (2005). Does education causes better health? A panel data analysis using school reform for identification, *Economic of Education Review* 22, 149-160.

Autor, D H., Katz, L.F., and Krueger, A. (1998). Computing inequality: Have computers changed the labor market? *Quarterly Journal of Economics,* 113: 1169-1213.

Autor, D H., Levy, F., and Murname, J.R. (2003). The skill content of recent technological change: An empirical exploration, *Quarterly Journal of Economics,* 118: 1279-1333.

Avery, R.B. (1977). Error components and seemingly unrelated regressions, *Econometrica*, 45, 199-209.

Beasley, T., (2008), Seemingly Unrelated Regression (SUR) Models as a Solution to the Path Analytic Models with Correlated Errors, *Multiple Regression Viewpoints*, 34, 1-5.

Becker, G. (1964). *Human capital: A theoretical and empirical analysis, with special reference to education,* Chicago, University of Chicago Press.

Berger, M., and Leigh, J. (1989). Schooling, self selection, and health, *Journal of Human Resources,* 24, 433-455.

Berhman, J.R., and Wolfe, B. (1989). Does more schooling make women better nourished and healthier? Adult sibling random and fixed effect estimates for Nicaragua, *Journal of Human Resources,* 15, 644-663.

Blackwell, J. (2005). Estimation and testing of fixed-effect panel-data systems, *The Stata Journal,* 5, 202-207.

Bound, J., and Johnson, G. (1992). Changes in structure of wages in the 1980s: An evaluation of alternative explanations, *American Economic Review,* 82: 371-392.

Bound, J., Jaeger, D.A., and Baker, R.M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak, *Journal of the American Statistical Association*, 90, 443-450.

Card, D., and DiNardo, J. (2002). Skill-biased technological change and rising wage inequality: Some problems and puzzles, *Journal of Labor Economics,* 20: 733-783.

Contoyannis, P., Jones, A. (2004). Socio-economics status, health, and lifestyle, *Journal of Health Economics*, 23, 965-995

Cowell, A. (2006). The relationship between education and health behavior: Some empirical eEvidence, *Health Economics* 15,124-146.

Currie, J., and Hyson, R. (1999). Is the impact of health stock cushioned by socioeconomic status? The case of low birthweight, *American Economic Review*, 89, 245-250.

Daldy, B., and Gibson, J. (2003). Have computers changed the New Zealand wage structure? Data on training. *New Zealand Journal of Industrial Relations,* 28: 13-21

Davies, M., Baer, D., Judd, J., Brown, E., Campbell, W., and Taylor, P. (2002). Effects of moderate slcohol intake on fasting insulin and glucose concentrations and insulin sensitivity in postmenopausal women, *The Journal of the American Medical Association, 287,* 2559-2562.

De Walque, D. (2007). Does education affect smoking behavior? Evidence using the Vietnam draft as an instrument for college education, *Journal of Health Economics,* 26, 877-895.

DiNardo, J., and Pichke, J. (1997). The return to computer use revisited: Have pencils changed the wage structure too? *Quarterly Journal of Economics,* 112: 291-303.

Dolton, P., and Makepeace, G. (2004). Computer use and earnings in Britain. *The Economic Journal,* 114, 114-129.

Doms, M., Dunne, T., and Troske, K.R. (1997). Workers, wages, and technology, *Quarterly Journal of Economics,* 112: 253-290.

Entrof, H., Golac, M., Kramarz, F. (1999). New technologies, wages, and worker selection, *Journal of Labor Economics,*17, 464-491.

Farell , P., and Fuch, V., (1982). Schooling and health: the cigarette connection, *Journal of Health Economics*, 1, 217-230

Fayissa, B., Nsiah, C., and Taddese, B. (2008). The impact of tourism on economic Growth and development in Africa, *Tourism Economics*, 14, 807-818.

Folland, S., Goodman, A., and Stano, M.(2001). *The economics of health and health care,* Prentice Hall, Upper Saddle River NJ.

Fuchs, V. (1974). *Who shall live: Health, economics, and social choice*, Basic Books Inc. New York.

Fuchs, V. (1982). Time preferences and health: An explanatory study, In V. Fuchs, Editor, Economics Aspect of Health, second NBER conference on health in Stanford, University of Chicago Press, 83-119.

Green, W. (2008). *Econometric analysis*, Pearson Education Inc., Upper Saddle River, NJ

Grimard, F, and Parent, D. (2003). Education and smoking: Were Vietnam War draft avoiders also more likely to avoid smoking? *Journal of Health Economics*, 26, 877-895.

Grossman, M. (1972). On the concept of health capital and the demand for health, *Journal of Political Economy*, 80, 223-225.

Grossman, M. (1973). Correlation between health and education, *NBER Working Paper* No. 22.

Grossman, M. (2005). Education and non-market outcomes, *NBER Working Paper*, No. 11582.

Joyce, T., Racine, A., and Mocan, N. (1992). The consequence and costs of maternal substance abuse in New York City: A pooled time-series, cross-section analysis, *Journal of Health Economics*, 11, 297-314.

Katz, L., and Murphy, K.M. (1992). Change in relative wages, 1963-1987: Supply and demand factors, *Quarterly Journal of Economics,* 107: 35-78.

Kenkel, D. (1991). Health behavior, health knowledge, and schooling, *Journal of Political Economy*, 99, 287-305.

Kenkel, D. (1995). Should you eat breakfast? Estimates from health production functions, *Health Economics*, 4, 15-29.

Krashinsky, H.A. (2004). Do marital status and computer usage really change the wage structure? *Journal of Human Resources,* 39: 774-791.

Krueger, A. (1993). How computers have changed the wage structure: Evidence from microdata, 1984-1989, *Quarterly Journal of Economics,* 108: 33-60.

Lemieux, T. (2006). Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skills? *American Economic Review,* 96: 461-498.

Lleras-Muney, A. (2005). The relationship between education and adult mortality in the U.S., *Review of Economic Studies*, 71, 189-221.

Maskara, P., Aggarwal, R. and Maskara, R. (2006). Access to computer technologies at home improves wages in the marketplace, *Journal of International Technology and Information Management,* 15: 67-77.

Mincer, J. (1991). Human capital, technology, and the wage structure: What do the time series show? NBER Working Paper No. 3581.

Mincer, J. (1958). Investment in human capital and personal income distribution, *Journal of Political Economy*, 66, 281-302.

Murphy, K., Welch, F. (1991). The structure of wages, *Quarterly Journal of Economics*, 106, 285-326.

Park, C., Kang, C. (2008). Does education induce healthy lifestyle? *Journal of Health Economics*, 27, 1516-1531.

Pabilonia, S., and Zoghi, C. (2005). Returning to the returns to computer use, *American Economic Review, Paper and Proceedings* (2005), 314-317.

Pigou, C. (1928). *A study in public finance*, Macmillan, London.

Register, C., William, D., Grimes, P. (2001). Adolescent drug use and educational attainment, *Eductional Economics*, 9, 1-18.

Rimm, E., Klasksy, A., Grobbee, D., and Stampfer, M. 919960. Review of moderate alcohol consumption and reduced risk of coronary heart disease: Is the effect due to beer, wine or spirits? *British Medical Journal, 312* (7033), 731-736.

Rosenzweig, M.R., and Schultz, T. (1983). Estimating a household production function: Heterogeneity, the demand for health inputs, and their effects on birth weight, *The Journal of Political Economy*, 91, 773-746.

Ross, C., and Wu, C. (1995). The links between education and health, *American Sociological Review*, 60, 719-745.

Sander, W. (1995). Schooling and quitting smoking, *Review of Economics and Statistics*, 77, 199-199.

Schultz, T. (1980). Nobel lecture: The economics of being poor, *Journal of Political Economy*, 88, 639-651.

Sillies, M. (2009). The casual effect of education on health: Evidence from United Kingdom, *Economic of Education Review*, 28, 122-128.

Smith, A. (1776). *An inquiry into the nature and causes of the wealth of nations, Book II: Of the nature, accumulation, and employment of stock*, Random House Inc.

Spitz-Oener, A. (2007). The return to pencil use revisited, IZA Discussion Paper No. 2729.

Staiger, D., and Stock, J. (1997). Instrument variable regression with weak instruments, *Econometrica* 65, 557-586.

Strauss, J., and Stock, J. (1997). Instrumental variables regressions with weak instruments, *Econometrica*, 65, 557-586.

Tashiro, S. (2004). The diffusion of computers and wages in the U.S.: Occupation and industry analysis, 1984-2001. Unpublished manuscript, Department of Economics, Rowan University.

U.S. Department of Health and Human Services and U.S. Department of Agriculture (DHHS). (2005). *Dietary Guidelines for Americans 2005*. Retrived from http://www health.gov/ dietaryguidelines/dga2005/doucument /html/chapter9.htm.

Weinberg, B.A. (2000). Computer use and demand for female workers. *Industrial and Labor Relations Review,* 53: 290-308.

Wolfe, B., and Berhman, M. (1987). Women's schooling and children's health: Are the effects robust with adult sibling control for the women's childhood background? *Journal of Health Economics,* 6, 239-254.

Wooldridge, J. (2002). Econometric analysis of cross section and panel data, The MIT Press, Cambridge, MA.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias, *Journal of the American Statistical Association* 57, 348-368.